Playing the field...

Light field technology enables production options

Light field technology enables capture of a richer scene representation with the potential to give more creative freedom to producers and artists, and to make the creation of visual effects easier. **Jean-Yves Guillemaut** and **Adrian Hilton** of the University of Surrey explain the current state of play in this exciting cutting-edge form of image capture.

Light field representation

Conventional cameras capture a 2D representation of a scene; they achieve very high spatial sampling (recording the light intensity at each scene point sampled by the sensor). This provides a photorealistic representation of the scene but all the creative decisions, such as shot composition, focus, depth of field, and so on, must be made at the point of capture and are baked into every frame.

It is not possible to alter these elements post-capture without going through a tedious post-production process and without risking degrading the realism of the shot. Besides, composition with computer-generated (CG) assets is not straightforward as conventional images are 2D and lack depth information.

In contrast, a light field is a 4D representation describing light rays travelling through space at all locations and in all directions (Levoy & Hanrahan, 1996; Gortler *et al.*, 1996). A light field captures both the spatial properties of the scene (the light intensity emitted by each scene point sampled by the camera, akin to conventional cameras) and the directional properties of the light emitted by the scene (the light intensity as a function of the direction in which light is reflected off surfaces in the scene).

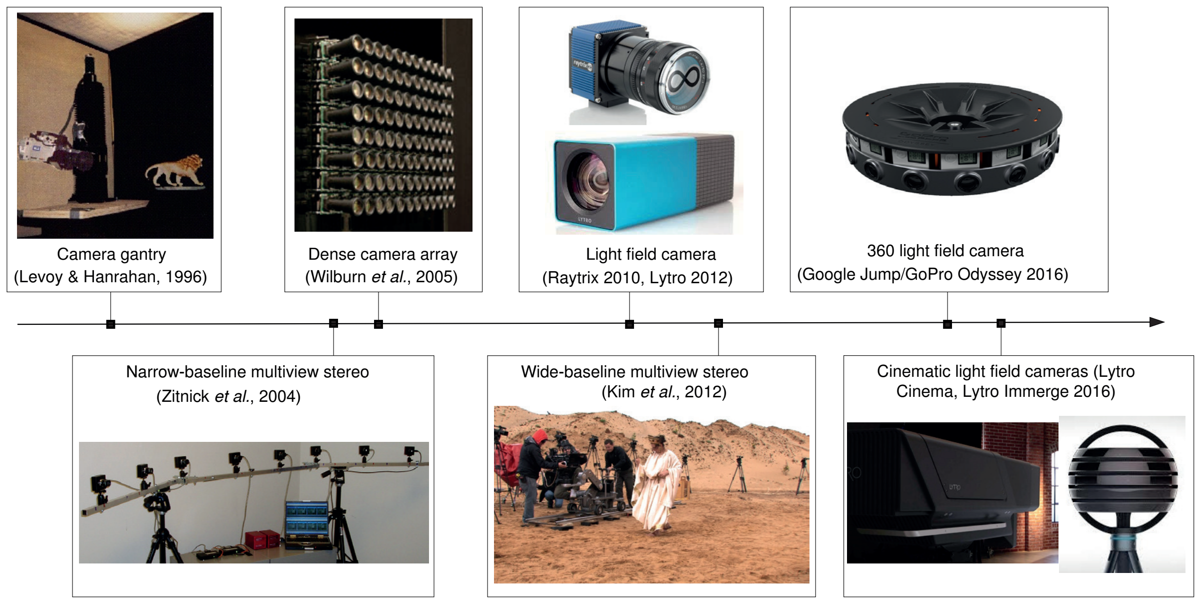
This richer representation provides added flexibility and creative control as some of the creative decisions that usually have to be made at the point of capture can now be postponed to post-production – without compromising on realism. For example, the director may decide to recompose the shot, alter the camera path or change the aperture and focus in post.

This new representation is also ideally tailored to immersive content production as it enables the viewpoint to be adjusted in real-time to match that of the user. Unlike 360° spherical capture, which is limited to head rotation and does not allow head movement to produce the parallax cues experienced in real-world scenes, light fields enable synthesis of views with correct parallax and photorealistic appearance.

Light field acquisition technology

The general principle of light field photography was originally pioneered in 1908 by Gabriel Lippmann under the name of integral photography; however, practical methods for light field acquisition have only just started to emerge in recent years (Figure 1).

Figure 1: Timeline of light field capture



Early approaches to capturing light fields performed a dense sampling of the 3D space by moving a camera around the scene. These approaches used either a camera mounted on a gantry (Levoy & Hanrahan, 1996) or more recently a handheld camera (Davis, Levoy & Durand, 2012). In these approaches, each frame effectively captures a 2D slice of the 4D light field. Due to the need to physically move the camera around the scene, these approaches are limited to static scenes. They use image-based rendering techniques to synthesise new views by interpolating the information from the captured views. As they do not explicitly capture the scene geometry, a large number of camera views (typically around 100) is required in order to densely sample the light field.

Multi-view stereo capture methods extend light field acquisition to the video domain by using a network of synchronised cameras distributed around the scene to simultaneously acquire 2D slices of the light field from multiple viewpoints. A 3D representation is then extracted using multi-view stereo reconstruction methods (Zitnick *et al.*, 2004; Starck & Hilton, 2007) and the 3D model is used as a geometric proxy to optimise blending of the different input views for photorealism (Buehler *et al.*, 2001).

Initial multi-camera light field setups were based on dense multi-camera arrays (i.e. a large number of cameras in close arrangement, as in Wilburn *et al.*, 2005). Recent research has explored the use of wide-baseline acquisition setups to reduce the number of required cameras (Starck & Hilton, 2007; Guillemaut & Hilton, 2011) and enable use in outdoor settings (Ballan *et al.*, 2010; Kim *et al.*, 2012, Mustafa *et al.*, 2015). These methods are more practical than their dense camera array counterparts but do not model all the view-dependent detail and there are limitations to the viewpoints that can be synthesised due to the sparse acquisition setup.

With the advent of computational photography and light field cameras (Ng *et al.*, 2005), it has become possible to capture a light field using a single sensor. Light field cameras (e.g. Raytrix, Lytro) insert a lenslet array between the main lens and the camera sensor to multiplex the 4D light field onto a 2D image sensor. Each micro lens separates the bundle of incoming light rays depending on their incoming directions, thus capturing the directional properties of incoming light at this particular location in the array. The complete lenslet array enables formation of an image which captures both spatial and directional information of the light field.

These cameras effectively trade some of the spatial resolution against angular resolution. Consequently, they require significantly more pixels than a conventional camera to achieve an equivalent final image resolution. Light field cameras are now being extended to compact stereoscopic 360° systems capable of producing photorealistic novel views and refocusing content (Google Jump, GoPro Odyssey). However, the viewpoint remains limited to a relatively small volume, only suitable for limited head motion. Very recently, Lytro has introduced the first professional cinematic light field camera called the Lytro Cinema, which was premiered at NAB 2016. They are also introducing the Lytro Immerge which is dedicated to cinematic virtual reality (VR) production.

A major advance enabled by light field capture is the ability to postpone the need to make hard decisions on shot composition, aperture and focus from capture time to render time. This removes the burden of having to get every element of the shot perfect at the time of capture. For example, it becomes unnecessary to perfectly track and maintain focus on small and rapidly moving objects since it is possible to refocus the shot at render time.

> A light field is a 4D representation describing light rays travelling through space at all locations and in all directions.

Some examples of visual effects that are enabled by this technology are:

- The light field representation enables extraction of depth information to facilitate visual effects. Depth extraction is made possible by the richer representation which samples the scene from a large number of viewpoints defined by the main lens aperture (in the case of a light field camera) or the different cameras in the setup (in the case of a camera array). The availability of depth as well as the multi-view appearance information allows the synthesis of new viewpoints by image-based rendering or texture mapping using the capture images, a process called novel view synthesis or free-viewpoint video.

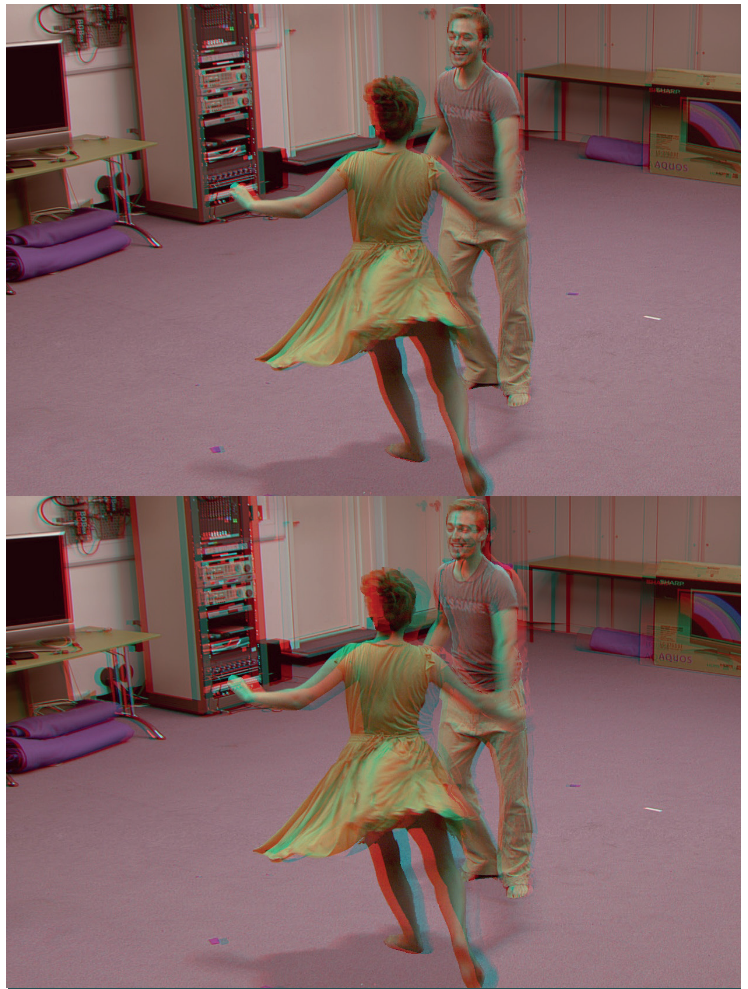


Figure 2: Example demonstrating the application to stereoscopic content production (stereoscopic images are rendered as anaglyph). The scene is captured using multiple cameras and stereoscopic content is produced in post by texture mapping the captured images onto a 3D model of the scene. This enables tailoring of the amount of depth to the display or user. The different images show the same scene rendering with varying amount of depth controlled by digitally changing the interaxial distance between the cameras at render time.
Images generated by University of Surrey as part of the TSB project i3DLive

Light field technology

> Some of the creative decisions that usually have to be made at the point of capture can now be postponed to post-production… the director may decide to recompose the shot, alter the camera path or change the aperture and focus in post.

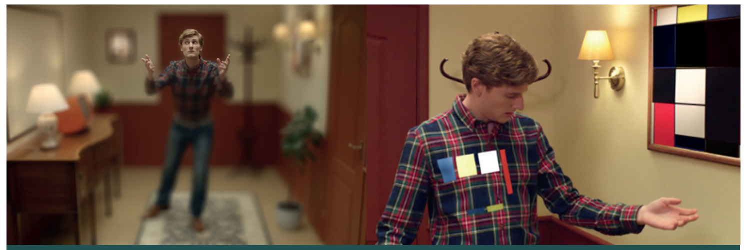


Figure 3: Examples demonstrating digital refocusing (left) and scene augmentation using computer-generated assets (right). In both cases, depth information (in this case captured using an additional depth sensor) is used to facilitate the generation of visual effects and increase creative control. *Images produced as part of the EU FP7 project SCENE*

- This can also be applied in the context of stereoscopic content production to control the amount of depth by digitally adjusting the interaxial distance at render time according to screen size or the personal preference of the user (Figure 2).
- The depth information opens up the possibility of achieving a range of other effects such as refocusing and changing the depth of field (Figure 3).
- It also facilitates traditional post-processing operations such as matting or compositing as the depth information can be used to constrain these tasks.

Producing photorealistic VR media experiences

Current VR production methods have significant limitations. 360° video delivers realistic imagery because it has been filmed in the real world, but the image is essentially flat; interactivity is limited (the scene is a single image, not made up of independent elements); most 360° video suffers from quality problems, especially on mobile VR platforms due to high levels of video compression; and people, although realistic, clearly look like video representations and so lack presence.

Real-time 'true VR' made with computer-generated imagery (CGI) is often sorely lacking in realism, particularly on mobile VR platforms. On desktop platforms, such as Oculus Rift and HTC Vive, locations can be produced to an acceptable level of photorealism. However, even high-level CGI character work (using photogrammetry and motion capture) lacks complete photorealism: the nuances of human performance, especially facial expressions and the movement of clothing, are often lost.

Figure 4: Four-camera rig used by MPC to capture *Goosebumps* VR demo

Current workarounds include filming actors on green screen with two- or four-camera rigs positioned in front of, or to one side of, the performer. However, the illusion breaks down when the user moves within the VR scene, revealing the secret of the trick – that the video is playing back on a 2D plane or a crude 3D mesh, which only looks correct from a restricted range of perspectives.

In contrast, light field technology directly captures a 4D representation and therefore naturally lends itself to immersive content production as it becomes possible to adjust the rendered content as a function of the user's position and orientation in the scene, while preserving the realism of the captured scene.

Commercially available 'outside-in' camera arrays (i.e. with cameras surrounding the scene and pointing inwards) are capable of producing content elements that can be viewed from multiple viewpoints, but they can only reproduce the surrounding environment from one point of view. Outside-in light fields are often recorded in a chromakey studio to isolate the actors/props, which can then be composited with a separately captured real background or CG background.

'Inside-out' camera arrays, such as commercially available compact stereoscopic 360° camera rigs, offer a limited range of potential viewpoints. More research is required to develop light field capture systems to support a larger range of viewpoints.

Light field challenges

Light fields have the potential to revolutionise the way we create and consume content; however, they also introduce a new set of challenges. Firstly, the light field representation is significantly larger than conventional image/video representations due to the increased dimensionality. This poses some challenges in terms of data storage and transmission. An important aspect to ensure acceptance of the technology will therefore be standardisation as well as the development of efficient compression techniques tailored specifically to light field data.

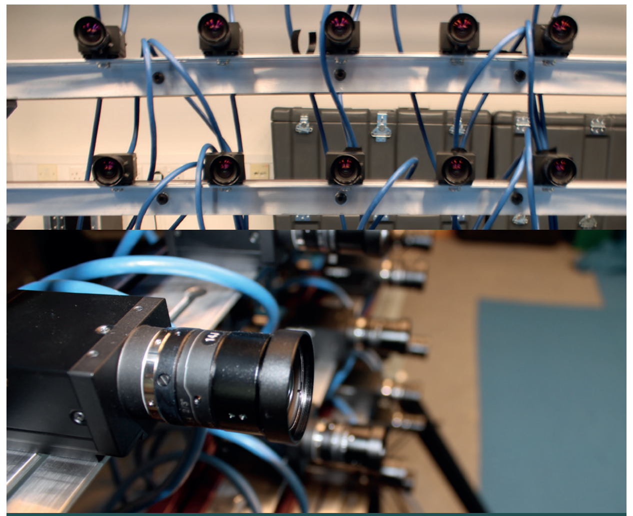


Figure 5: The light field capture setup for immersive VR production developed at the University of Surrey as part of the InnovateUK project ALIVE. This consists of an array of 25 cameras whose positions can be adjusted according to the scene dimensions and expected range of motion of the end-user.

Another challenging aspect relates to the development of creative tools dedicated to the light field editing process. There is currently a lack of software tools for light field post-production. Finally, giving the user the ability to freely explore and move in a scene introduces new challenges in terms of storytelling and narratives for VR content production.

Increasing creative freedom and flexibility

Whether based on multiple camera setups or microlens array technology, light fields now offer content-makers a whole new degree of flexibility and freedom to create visual effects.

References

Ballan L *et al.* (2010). *Unstructured video-based rendering: interactive exploration of casually captured videos.* In: ACM SIGGRAPH.

Buehler C *et al.* (2001). *Unstructured lumigraph rendering.* In: Proc. SIGGRAPH, 425–32.

Davis A, Levoy M, Durand F (2012). Unstructured light fields. *Computer Graphics Forum*, **31**(2):305–14.

Gortler SJ *et al.* (1996). *The lumigraph.* In: ACM SIGGRAPH, 43–54.

Guillemaut J-Y, Hilton A (2011). Joint multi-layer segmentation and reconstruction for free-viewpoint vdeo applications. *International Journal of Computer Vision*, **93**(1):73–100.

Kim H *et al.* (2012). Outdoor dynamic 3-D scene reconstruction. *IEEE Transactions on Circuits and Systems for Video Technology*, **22**(11):1611–22.

Levoy M, Hanrahan P (1996). *Light field rendering.* In: ACM SIGGRAPH, 31–42.

Mustafa A *et al.* (2015). *General dynamic scene reconstruction from multiple view video.* In: Proc. ICCV, 900–8.

Ng R *et al.* (2005). *Light field photography with a hand-held plenoptic camera.* Stanford University Computer Science Tech Report CSTR 2005-02.

Starck J, Hilton A (2007). Surface capture for performance-based animation. *IEEE Computer Graphics and Applications*, **27**(3):21–31.

Wilburn B *et al.* (2005). *High performance imaging using large camera arrays.* In: ACM SIGGRAPH, 765–76.

Zitnick CL *et al.* (2004). High-quality video view interpolation using a layered representation. *ACM Transactions on Graphics (TOG)*, **23**(3).

Fact File

Contact the authors:
Jean-Yves Guillemaut: j.guillemaut@surrey.ac.uk

Adrian Hilton: a.hilton@surrey.ac.uk

The Centre for Vision, Speech and Signal Processing (CVSSP) is one of the major research centres of the Faculty of Engineering and Physical Sciences at the University of Surrey.

Its aim is to advance the state of the art in multimedia signal processing and computer vision, with a focus on image, video and audio applications. CVSSP, directed by Prof Adrian Hilton, is one of the largest research centres in the UK focusing on signal processing and interpretation. It has over 120 members comprising academic and support staff, postdoctoral research fellows, visitors and PhD students. CVSSP research income is in excess of £18.5m, with about 90% resourced by grants from EPSRC, European Commission, InnovateUK and industry.